

# Gravitationally unstable gaseous disks of flat galaxies<sup>★</sup>

E. Griv

Department of Physics, Ben-Gurion University of the Negev, PO Box 653, Beer-Sheva 84105, Israel  
e-mail: griv@bgu.ac.il

Received 9 June 2005 / Accepted 7 November 2005

## ABSTRACT

**Context.** The dynamics of a self-gravitating gaseous subsystem of a disk galaxy is considered analytically, using a local WKB approximation in the radial direction. The simplified model of a galaxy is used in which stars (and a dark matter, if it exists at all) do not participate in the disk collective oscillations and just form a background charge.

**Aims.** For the first time in galactic dynamics, an equation is derived to describe the torque that results from the buildup of gravitational Jeans-type instability of nonaxisymmetric gravity perturbations (e.g., those produced by a spontaneous disturbance or, in rare cases, a satellite system).

**Methods.** The torque serves to redistribute the angular momentum of the inhomogeneous disk by growing (that is, unstable) nonaxisymmetric density waves in such a way that a fraction of the mass (eventually residing in the central parts) retains only a minor fraction of the angular momentum, most of the latter being deposited in extended outer regions of the system.

**Results.** The outward transfer of angular momentum (and the “heating”) bring the disk toward stability against all small perturbations, including the most unstable nonaxisymmetric ones.

**Key words.** hydrodynamics – instabilities – turbulence – waves – ISM: general – galaxies: kinematics and dynamics

## 1. Introduction

Gas-rich spiral galaxies have been discussed in terms of the density wave theory, which views the spiral arms as wave patterns (Lin et al. 1969; Shu 1970; Lin & Lau 1979). Lin and Shu were the first to calculate the collective behavior of the stars and gas in a satisfactory formulation. We have approached the study of spiral galaxies by considering the dynamics of a gravitating disk of stars. It has been suggested that a natural explanation of the origin of the density waves can be obtained if one assumes them to be unstable normal modes of collective oscillations of galaxies. The basic quasilinear theory of gravitational Jeans-type instability for pure stellar disks has been developed by Griv et al. (2002) and Griv & Gedalin (2004). Here, the galactic gaseous disk is studied. The simplified model of a galaxy is used in which stars (and a dark matter, if it exists at all) do not participate in the disk oscillations and just form a background charge. (The bulk of the total optical mass, probably  $\gtrsim 90\%$ , in the Milky Way and other flat galaxies is in stars. HIPPARCOS data (Haywood et al. 1997) indicate only a moderate contribution of unseen matter to the local potential.) A Lagrangian description of the motion of a gas element under the influence of a perturbed gravity field is used, looking for time-dependent waves which propagate in a nonuniformly rotating, inhomogeneous, two-dimensional disk. Such a fluid model can also be

used as a zeroth-order approximation of a galaxy composed primarily of stars (Lin & Lau 1979). The approximation of an infinitesimally thin disk is a valid approximation if one considers perturbations with a wavelength greater than the disk thickness  $2h$  (Shu 1984).

There have been a number of prior studies in the literature which have investigated the growth of instabilities in gravitating gaseous disks. Lau & Bertin (1978), Lin & Lau (1979), Morozov (1985) established the difference in the stability boundary between axisymmetric and nonaxisymmetric disks by theoretical arguments. Goldreich & Tremaine (1979) calculated the linear response of a differentially rotating disk to a rigidly rotating external potential. It was shown that the external potential exerts torques on the disk only at the spatially limited resonances. In gravitationally unstable systems disk motions caused by self-gravity become perhaps dominant sources of angular momentum transport (Paczynski 1978; Lin & Pringle 1990). Angular momentum transport enhanced by the gravitational instability was investigated in gaseous accretion disks by  $N$ -body simulations with self-gravitating particles (Tomley et al. 1991) and hydrodynamical simulations (Laughlin & Różyczka 1996; Laughlin et al. 1998; Gammie 2001; Pickett et al. 2003). Simulations have already demonstrated that the disk is prone to a series of  $m = 1-4$  spiral modes. The gravitational torques induced by these  $m \neq 0$  structures elicit material transport of angular momentum and mass through the disk, readjusting the density profile toward more

<sup>★</sup> Appendices are only available in electronic form at <http://www.edpsciences.org>

stable configuration. As angular momentum is transferred outward, inner material falls to the central body and outer material migrates outward. Papaloizou & Lin (1995) have reviewed theoretical investigations regarding the transport of angular momentum in accretion disks. It has been *suggested* that the angular momentum can be transported by torques resulting from the presence of unstable modes in disks. However, no equations explicitly describing this transport process have been obtained.

Summarizing, gravitational instabilities as a mechanism for the transfer of angular momentum (and “heating”) in gaseous disks have been studied numerically. However, the transport of angular momentum, while fairly well understood within the computer models, remains a significant challenge for theoretical models of galaxies.

My significant contribution is just a hydrodynamic theory derivation of results obtained before in simulations discussed above. In the spirit of Lau & Bertin (1978) and Lin & Lau (1979), I carry out a second order Lin-Shu asymptotic approximation; thus azimuthal gradients of perturbed quantities are included in addition to the local gradients of equilibrium parameters and the differential rotation. In addition, a different approach from the rest is taken: to consider the complex conjugates of perturbed quantities. Without *all* these terms, the model fails to describe the exchange of angular momentum and the mass redistribution in the wave-gas system. I find that spiral density waves can be self-excited in the main domain of the disk between the inner and outer Lindblad resonances via the dynamical Jeans instability (gravitational collapse) in a nonresonant wave-“fluid” interaction, and can effectively transport the angular momentum and mass in the disk. I show that the hydrodynamic turbulence due to nonresonant gravitational instability transports the angular momentum, giving the disk a source of internal viscosity (the angular momentum flux)<sup>1</sup>. Also, instabilities heat the disk up on the fast dynamical timescale, bringing it toward stability. Another important result is that the critical sound speed for disk stability is approximately equal to 2 times the familiar Safronov-Toomre value.

It should be made clear right from the start that Lau & Bertin (1978) and Lin & Lau (1979) already carried out the second order Lin-Shu approximation in their analyses. To obtain the general picture of disk stability, I repeat some results of the Lau, Bertin, and Lin theories by using a slightly different mathematical technique. But even in this case some new results are obtained, e.g., an expression for the pattern speed  $\Omega_p$  of Jeans-unstable waves.

## 2. Perturbed motion

The time dependent surface density  $\sigma(\mathbf{r}, t)$  of the gaseous subsystem is split up into a basic and a developing (perturbation) part,  $\sigma = \sigma_0(r) + \sigma_1(\mathbf{r}, t)$  and  $|\sigma_1/\sigma_0| \ll 1$  for all  $\mathbf{r}$  and  $t$ ,

<sup>1</sup> That a turbulent viscosity could be one agent for angular momentum transport in a differentially rotating medium has been postulated (Shakura & Sunyaev 1973; Lynden-Bell & Pringle 1974; Pringle 1981). Another agent is stable self-gravitational waves, including those generated by resonant interactions (Goldreich & Tremaine 1979). The hydrodynamic turbulence that may arise as a result of gravitational instability is related to stochastic motions of gas elements.

where  $(r, \varphi, z)$  are the galactocentric cylindrical coordinates and the axis of the disk rotation is taken oriented along the  $z$ -axis. The total gravitational potential  $\Phi(\mathbf{r}, t)$  and the gaseous pressure  $P(\mathbf{r}, t)$  are also of this form. These quantities  $\sigma$ ,  $\Phi$ ,  $P$  are then substituted into the equations of motion of the gas element, the continuity equation, the Poisson equation, and the second order terms of the order of  $\sigma_1^2$ ,  $\Phi_1^2$ ,  $P_1^2$  may be neglected with respect to the first order terms.

The resultant equations of motion are cyclic in the variables  $t$  and  $\varphi$ , and hence by applying the standard lowest (or local) WKB method (Krall & Trivelpiece 1986) one may seek solutions in the form of normal modes:

$$\sigma_1(\mathbf{r}, t) = \sum_k \tilde{\sigma}_k \exp(ik_r r + im\varphi - i\omega_k t) + \text{c.c.}, \quad (1)$$

where  $\tilde{\sigma}_k$  is the real amplitude,  $k_r$  is the real radial wavenumber,  $m$  is the nonnegative azimuthal mode number,  $\omega_k = \Re\omega_k + i\Im\omega_k$  is the complex wavefrequency, suffixes  $k$  denote the  $k$ th Fourier component, and “c.c.” means the complex conjugate. The varying part of  $\sigma_1$  is absorbed in its phase. Evidently  $\sigma_1$  is a periodic function of  $\varphi$ , and hence  $m$  must be an integer. The criteria for stability differ for each  $m$ , and must be determined by a detailed analysis. The assumption that  $\tilde{\sigma}_k$  has a weak spatial dependence corresponds to the quasiclassical approximation in quantum mechanics and to the approximation of geometrical acoustics in the propagation of light in an inhomogeneous medium (Landau & Lifshits 1987). The imaginary part of  $\omega_k$  corresponds to a growth ( $\Im\omega_k > 0$ ) or decay ( $\Im\omega_k < 0$ ) of the components in time,  $\sigma_1 \propto \exp(\Im\omega_k t)$ , and the real part to a rotation with angular velocity  $\Omega_p = \Re\omega_k/m$ . When  $\Im\omega_k > 0$ , the medium transfers its energy to the growing wave. A galaxy is considered to be a superposition of different oscillation modes. A disturbance in the disk will grow until it is limited by some nonlinear effect. In the WKB method, the radial wavenumber is presumed to be of the form  $k_r(r) = \mathcal{A}\Psi(r)$ , where  $\mathcal{A}$  is a large parameter and  $\Psi(r)$  is a slowly varying function of the radial distance  $r$ , i.e.,  $d \ln k_r / d \ln r = O(1)$ . In the local approximation the wave is considered plane, i.e., all derivatives of  $k_r(r)$  and  $\tilde{\sigma}(r)$  are fully neglected. When calculating the terms of higher order one can simultaneously solve the wave equations with any desired degree of accuracy. It is assumed here, by considering the two-dimensional system, that  $\mathbf{k}(r)$  is perpendicular to the rotation axis, i.e.,  $k^2$  is given by  $k^2 = k_r^2 + k_\varphi^2 \equiv k_r^2 + m^2/r^2$  and  $k_\varphi = m/r$  is the azimuthal wavenumber. It is remarkable that the results of WKB theory are in agreement with the simulations (Mayer et al. 2004). In the linear theory, in Eq. (1) one can select one of the harmonics:  $\sigma_1(\mathbf{r}, t) = \tilde{\sigma} \exp(ik_r r + im\varphi - i\omega t) + \text{c.c.}$  (cf. Landau & Lifshitz 1987, p. 261). The potential  $\Phi_1(\mathbf{r}, t)$  and the pressure  $P_1(\mathbf{r}, t)$  are also of this form. The solution in such a form represents a plane spiral wave with  $m$  arms or a ring ( $m = 0$ ). With  $\varphi$  increasing in the rotation direction, we have  $k_r > 0$  for trailing patterns, and  $k_r < 0$  for leading ones. With  $m = 0$ , we have the density waves in the form of concentric rings that propagate away from the center when  $k_r > 0$ , or toward the center when  $k_r < 0$ . A weakly inhomogeneous medium is considered:  $|k_r r| \gg |k_r L| \gg 1$ , where  $|L|$  is the radial scale of inhomogeneity.

The disk is subject to the equations of motion. The linearized equations of two-dimensional motion of the gas element in the frame of reference rotating with angular velocity  $\Omega$  at the reference position  $r_0$  are

$$\frac{dv_r}{dt} - 2\Omega v_\varphi + 2r_0 r_1 \Omega \frac{d\Omega}{dr} = -\frac{\partial\Phi_1}{\partial r} - \frac{c_s^2}{\sigma_0} \frac{\partial\sigma_1}{\partial r}, \quad (2)$$

$$\frac{dv_\varphi}{dt} + 2\Omega v_r = -\frac{1}{r_0} \frac{\partial\Phi_1}{\partial\varphi} - \frac{c_s^2}{r_0\sigma_0} \frac{\partial\sigma_1}{\partial\varphi}, \quad (3)$$

where  $v_r$  and  $v_\varphi$  are the radial and azimuthal peculiar velocities,  $\Omega = \Omega(r_0)$ ,  $c_s = (\partial P/\partial\sigma)^{1/2}$  is the speed of sound,  $r_1$  is a perturbation,  $r = r_0 + r_1$ ,  $|r_1/r_0| \ll 1$ , and the expansion  $P = P_0 + \sigma_1 (\partial P/\partial\sigma)_0$  is used. The influence of stars and halo matter enters through  $\Omega(r)$ .

Equations (2) and (3) must be solved simultaneously with the linearized continuity equation

$$\sigma_1 \approx - \int_{-\infty}^t \left( \sigma_0 \frac{\partial}{\partial r} v_r + v_r \frac{\partial\sigma_0}{\partial r} + \frac{\sigma_0}{r} \frac{\partial}{\partial\varphi} v_\varphi \right) dt', \quad (4)$$

where  $v_r$  and  $v_\varphi$  are now the perturbed (“forced”) velocities, the subscript of  $r_0$  is dropped as we are considering linearized quantities, and the relatively small term  $\sigma_0 v_r/r$  is omitted, i.e., the curvature effect is neglected. This is a valid approximation if  $|k_r r| \gg 1$  (Lin & Lau 1979). In the WKB approximation in the radial direction, the second term  $\propto \partial\sigma_0/\partial r$  and the third term  $\propto \partial v_\varphi/\partial\varphi$  in the bracket are much smaller than the first one  $\propto \partial v_r/\partial r$ .

For such a form of  $\Phi_1$  (Eq. (1)) the particular solution of the system of Eqs. (2)–(3) is

$$v_r = \frac{\aleph}{\omega_*^2 - \kappa^2} \left( \omega_* k_r + i \frac{2m\Omega}{r} \right) + \text{c.c.}, \quad (5)$$

$$v_\varphi = \frac{\aleph}{\omega_*^2 - \kappa^2} \left( \frac{4\Omega^2 - \kappa^2 + \omega_*^2 m}{\omega_*} \frac{m}{r} - i 2\Omega k_r \right) + \text{c.c.}, \quad (6)$$

where  $\aleph = \Phi_1 + c_s^2 \sigma_1/\sigma_0$ ,  $\omega_* = \omega - m\Omega$  is the Doppler-shifted (in a rotating frame) wavefrequency,  $\kappa = 2\Omega [1 + (r/2\Omega)(d\Omega/dr)]^{1/2}$  is the epicyclic frequency, in galaxies  $d\Omega/dr < 0$ ,  $\kappa \gtrsim \Omega$ , and  $\omega_* \neq 0$ ,  $\omega_*^2 - \kappa^2 \neq 0$  (Lin & Lau 1979; Griv et al. 1999). The solutions (5) and (6) describe the small forced velocities of the gas element in the radial and azimuthal directions under the action of the gravity perturbation,  $|v_r|$  and  $|v_\varphi| \ll r\Omega$ . If both the expressions are inserted into Eq. (4), a relation between the fluctuations in density and potential  $\sigma_1 = \sigma_1(r, \Phi_1)$  is obtained. A second relation of this kind is supplied by the linearized Poisson equation  $\Delta\Phi_1 = 4\pi G\sigma_1\delta(z)$ .

As is seen, the present theory suggests some systematic radial and azimuthal motions of the gas element distributed in the form of a spiral flow field which is a correction to the basic equilibrium motion,  $r\Omega^2 = \partial\Phi_0/\partial r + (c_s^2/\sigma_0)(\partial\sigma_0/\partial r)$ . In this equation, the term  $\propto c_s^2$  is a correction. Equilibrium is established in a simple manner in such a disk, i.e., it is governed mainly by the balance between the centrifugal and gravitational forces. In the unperturbed motion the gas elements describe circular orbits. It is this equilibrium model of a galaxy that is to be examined for stability against perturbations in the present

investigation. To repeat myself, the total potential  $\Phi_0(r)$  is produced mainly by stars (and dark matter).

### 3. Oscillation spectrum

#### 3.1. Perturbed surface density

Substituting the expressions for velocities (5)–(6) into the linearized continuity equation (4) and carrying out the integration in Eq. (4) over  $t$ , it is straightforward to show that the perturbed density can be expressed as

$$\sigma_1 = \frac{\sigma_0 \aleph}{\omega_*^2 - \kappa^2} \left( k_r^2 + \frac{\gamma m^2}{\omega_*^2 r^2} + \frac{2\Omega m}{\omega_* rL} \right) + \text{c.c.}, \quad (7)$$

where  $\sigma_1(t \rightarrow -\infty) = 0$ , so by considering only growing perturbations ( $\Im\omega_* > 0$ ) the effects of the initial conditions are neglected,  $\gamma = 4\Omega^2 - \kappa^2 + \omega_*^2$ , and

$$|L| = \left| \left( \frac{d}{dr} \right) \ln \left( \sigma_0 \Omega \kappa^{-2} \right) \right|^{-1}$$

is the radial scale of spatial inhomogeneity. Equation (7) contains the second-order term in the bracket:

$$\frac{4\Omega^2 - \kappa^2 + \omega_*^2 m^2}{\omega_*^2 r^2} \ll k_r^2$$

as well as the term  $\propto mL^{-1}$ , which play a crucial role in modifying Safronov-Toomre’s stability criterion from  $Q \equiv c_s/c_T \geq 1$  to  $Q \gtrsim 2\Omega/\kappa \approx 2$  (see Sect. 3.3 below for an explanation). These additional terms come from second order theory:  $m$ -dependent terms in the exponential factor of Eq. (1), the second and the third terms in the bracket of Eq. (4). The results of Lau & Bertin (1978) and Lin & Lau (1979) are the same, with the above terms (the terms  $\propto m^2/r^2$ ,  $T_1$ , and  $d \ln(\kappa^2/\sigma_0\Omega)/d \ln r$  of Lin & Lau’s (1979) Eq. (D12)). Equation (7) describes the response of a differentially rotating ( $4\Omega^2 > \kappa^2$ ) and inhomogeneous fluid disk to gravity perturbations. In contrast to Griv et al. (1999), only the most important low-frequency ( $|\omega_*|^2 \lesssim \kappa^2$ ) perturbations developing in the  $z = 0$  plane between the inner and outer Lindblad resonances are considered.

In Eq. (7), the term  $\propto L^{-1}$  corresponds to a wave-fluid resonance  $\omega_* \rightarrow 0$ , and sufficiently far from the resonance this term may be omitted (Lin & Lau 1979). However, as it will be shown in Sect. 4 below the existence of inhomogeneity is critically important for the exchange of angular momentum in the wave-gas system, and therefore the term  $\propto L^{-1}$  must be retained. The resonance  $\omega_* = 0$  has been studied by Lovelace & Hohlfield (1978).

#### 3.2. Generalized dispersion relation

On the other hand, the asymptotic ( $k_r^2 \gg k_\varphi^2$ ) solution of the Poisson equation is

$$\sigma_1(r) = -|k|\Phi_1/2\pi G + \text{c.c.}, \quad (8)$$

where the factor  $\frac{1}{4}$  is neglected compared with  $m^2$  (Bertin & Mark 1978). Solution (8) determines the perturbed density required to support the perturbed potential up to second order in the Lin-Shu asymptotic approximation (Bertin & Mark 1978;

Lau & Bertin 1978; Lin & Lau 1979). In the first approximation ( $m = 0$ ), from Eq. (8) it follows the old Lin-Shu result (Lin et al. 1969): maxima of the surface density correspond to minima of the potential,  $\sigma_1 = -|k_r|\Phi_1/2\pi G + \text{c.c.}$ . Note that the condition  $|m/k_r r|^2 \ll 1$  used by Bertin & Mark (1978) and Lin & Lau (1979) to obtain the asymptotic second-order solution (8) of the Poisson equation is equivalent to the requirement  $|\tan \psi| \lesssim 1$ , where the pitch angle  $\psi = \arctan(m/rk_r)$  is the angle between the direction of the wave front and the tangent to the circular orbit of the gas element, with “ $\lesssim$ ” in place of “ $\ll$ ” in contrast to the first-order approximation (Lin et al. 1969; Binney & Tremaine 1987, Sect. 6.2 therein) of the Lin-Shu theory. The asymptotic Lin-Shu approximation used here is actually the WKB approximation, which is tight winding approximation for small  $m$  (Griv et al. 1999, 2002).

Equating the density  $\sigma_1$  (Eq. (7)) to the perturbed density given by Eq. (8), one obtains the generalized Lin-Shu dispersion relation  $\omega_* = \omega_*(r, k)$ , which determines the spectrum of collective oscillations of fluid disks,

$$1 = \frac{k^2 c_s^2 - 2\pi G \sigma_0 |k|}{\omega_*^2 - \kappa^2} \left( \frac{k_r^2}{k^2} + \frac{\gamma}{\omega_*^2} \frac{m^2}{k^2 r^2} + \frac{2\Omega}{\omega_*} \frac{m}{k^2 r L} \right), \quad (9)$$

where  $2\pi G \sigma_0 |k| \gg k^2 c_s^2$ , the second term in the bracket proportional to  $m^2/r^2$  is a small correction (this is because  $4\Omega^2 \sim \kappa^2$  and  $m^2/k_r^2 r^2 \ll 1$ ), and  $|\omega_*^2| \lesssim \kappa^2$ . The dispersion relation (9) is of the order of  $\omega_*^2$  and is consistent with the dispersion relation derived by Montenegro et al. (1999, their Eq. (20)). Neglecting the effects of spatial inhomogeneity ( $|L| \rightarrow \infty$ ), in Griv et al. (1999), the generalized dispersion relation is given by Eq. (37), which is of the order of  $\omega_*^4$  (cf. Eq. (9) above) and is consistent with the dispersion relation obtained by Bertin et al. (1989, their Eq. (3.1)), which is of the order of  $\omega_*^4$  also.

Equation (9) is complicated: it is nonlinear in  $\omega_*$ . In order to deal with the most interesting oscillation types, let us consider various limiting cases of perturbations described by some simplified variations of Eq. (9). For instance, one can solve Eq. (9) by successive approximations (cf. Lifshitz & Pitaevskii 1981, p. 115). In the low-frequency ( $\omega_*^2 \ll \kappa^2$ ) and local WKB approximations we are explored in Eq. (9), the terms that describe tangential forces are assumed to be small in comparison with other terms. By neglecting the small terms  $\propto m^2/r^2$  and  $m/rL$ , one obtains, in the first approximation, the standard Lin-Shu dispersion relation for axisymmetric waves

$$\omega^2 = \kappa^2 - 2\pi G \sigma_0 |k_r| + k_r^2 c_s^2 \quad (10)$$

(Lin et al. 1969; Binney & Tremaine 1987, p. 352). From this equation, the disk is Jeans-unstable ( $\omega^2 < 0$ ) to axisymmetric perturbations if

$$c_s < c_T, \quad (11)$$

where  $c_T = \pi G \sigma_0 / \kappa$  is the Safronov-Toomre (Safronov 1960; Toomre 1964) critical sound speed to suppress the instability of only axisymmetric  $m = 0$  perturbations. On the other hand, writing the sound speed as  $c_s = \sqrt{\mathcal{R}T}$ , the Safronov-Toomre instability condition (11) reads  $T < \pi^2 G^2 \sigma_0^2 / \mathcal{R} \kappa^2$ , where  $T$  is the local gas temperature.

In the next approximation, in Eq. (9), considering the Safronov-Toomre stable disk ( $\omega^2 > 0$ , or  $c_s > c_T$ ), in the small term  $\propto m^2/r^2$  one can replace  $\omega_*^2$  by  $\kappa^2$ . As a result, the simplified generalized dispersion relation is obtained:

$$\omega_*^3 - \omega_* \omega_J^2 + 4\pi G \sigma_0 \Omega (m/r|k|L) = 0, \quad (12)$$

where  $|\omega_*| \lesssim \kappa$ ,

$$\omega_J^2 = \kappa^2 - 2\pi G \sigma_0 (k_*^2/|k|) + k_*^2 c_s^2 \quad (13)$$

is the squared Jeans frequency,  $k = \sqrt{k_r^2 + m^2/r^2}$ , and  $k_*^2 = k^2 \{1 + [(2\Omega/\kappa)^2 - 1] \sin^2 \psi\}$  is the squared effective wavenumber. Equation (12) differs from the Lin-Shu dispersion relation (10) by the appearance of the total  $k$  and effective  $k_*$  wavenumbers, which originate from the consideration of the nonaxisymmetrical modes  $\propto m$ , and by the factor  $\propto L^{-1}$ , which originate from the consideration of the effects of inhomogeneity (Lin & Lau 1979; Morozov 1985; Griv et al. 2002). Note that Lin et al. (1969), Shu (1970), as well as Toomre (1964, 1977) and Mark (1977), allowed for a departure from axial symmetry of the perturbations only partially by introducing a Doppler-shifted wavefrequency  $\omega_* = \omega - m\Omega$  but omitting all other  $m$ -dependent terms in the exponential factor of Eq. (1); that is, in fact they ignored the effect of the finite spiral inclination. Apparently, Lynden-Bell & Kalnajs (1972, their Eq. (A11)) first obtained the Lin-Shu-type dispersion relation for open waves propagating in a homogeneous disk. Simplified generalized dispersion relation (12) for low-frequency  $|\omega_*|^2 < \kappa^2$  perturbations we are interested in can be easily obtained from Eqs. (D12) (actually, in Eq. (D12) it should be  $T_1/(1-\nu^2)$  instead of  $T_1$ . See Montenegro et al. 1989) and (D14) of Lin & Lau (1979) by ignoring the “out-of-phase” term  $ik_r A$  and using the expansion  $2\Omega/\kappa \approx 1 - (r/4\Omega)(d\Omega/dr)$ . Also, the dispersion relation (12) becomes identical with the dispersion relation (38) of Griv et al. (1999) (obtained by using the microscopic kinetic-theory equations) by neglecting the effects of Landau-type wave-star resonances.

### 3.3. Modified Safronov–Toomre stability criterion

The dispersion relation (12) has three roots that describe three branches of oscillations: two ordinary Jeans branches modified by the inhomogeneity and a gradient one. Let us investigate these branches of oscillations following closely Morozov (1981, 1985) and Griv et al. (2002). From Eq. (12) in the most important high-frequency range

$$|\omega_*|^3 \sim |\omega_J|^3 \gg 4\pi G \sigma_0 \Omega (m/r|k|L), \quad (14)$$

we determine the dispersion law for the Jeans branch:

$$\omega_{*1,2} \approx \pm p |\omega_J| - 2\pi G \sigma_0 (\Omega/\omega_J^2) (m/r|k|L), \quad (15)$$

where  $p = 1$  for gravity-stable perturbations with  $\omega_*^2 \approx \omega_J^2 > 0$ ,  $p = i$  for gravity-unstable perturbations with  $\omega_*^2 \approx \omega_J^2 < 0$ , and the term involving  $L^{-1}$  is the small correction. Equation (15) determines the spectrum of oscillations. Accordingly, an inhomogeneity will not influence the stability condition of Jeans modes.

As is seen from Eq. (15), in contrast with the propagating or convective instability, the  $m = 0$  and/or  $L^{-1} \rightarrow 0$  instabilities give  $\Im\omega_{*1,2} > 0$  with  $\Re\omega_{*1,2} = 0$  and therefore are nonpropagating or absolute instabilities (Lifshitz & Pitaevskii 1981, Sect. 6 therein). The equilibrium parameters of the disk and the mode number  $m$  determine the pattern speed of Jeans-unstable ( $\Im\omega_{*1,2} > 0$ ) spiral perturbations (in a rotating frame)  $\Omega_p = \Re\omega_{*1,2}/m$  or

$$\Omega_p = -\frac{2\pi G\sigma_0}{\omega_J^2} \frac{\Omega}{r|k|L} \sim \frac{\Omega}{rLk^2}, \quad (16)$$

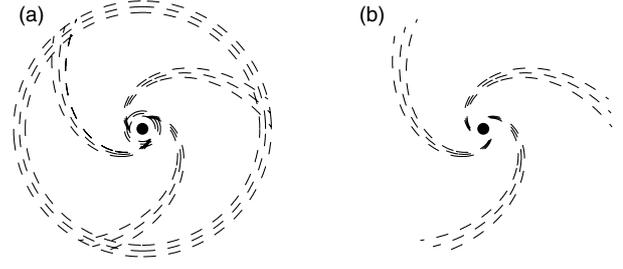
where  $\omega_J^2 < 0$  and  $r|L|k^2 \gg 1$ . Because in Eq. (16)  $r|L|k^2 \gg 1$ , the typical pattern speeds of spiral structures in Jeans-unstable disks are only a small fraction of some average angular velocity  $\Omega_{\text{ave}}$ . The theory states that in spatially homogeneous ( $|L| \rightarrow \infty$ ) disks  $\Omega_p = 0$ . This is a new result: in the standard density wave theory the pattern speed of spiral structure of a galaxy  $\Omega_p$  is a free parameter. In an inhomogeneous disk  $\Re\omega_* \neq 0$  and, therefore,  $\Omega_p \neq 0$ , and hence in this case the concept of slowly ( $|\Re\omega_*/\Im\omega_*| \ll 1$ ) propagating spiral density waves is meaningful. Also, because  $\Omega_p$  does not depend on  $m$ , each Fourier component of a nonaxisymmetric gravity perturbation in the inhomogeneous system will rotate with the same constant angular velocity.

From Eq. (13), the disk is Jeans-unstable ( $\omega_J^2 < 0$ ) to both axisymmetric (radial) and nonaxisymmetric (spiral) perturbations if  $c_s < c_T^2$ . Thus, the above results suggest that disks with Toomre's  $Q \equiv c_s/c_T$  parameter  $< 1$  break up into fragments of preferred mass  $M_{\text{frag}} \sim \lambda_{\text{crit}}^2 \sigma_0 = 4c_s^4/G^2\sigma_0$ . Similar results hold for a Safronov-Toomre unstable disk of finite thickness with the wavenumber  $k_{\text{crit}} = 2\pi/\lambda_{\text{crit}}$  reduced by a factor of 2 (Morozov 1981).

Nonaxisymmetric ( $m \neq 0$ ) instabilities in a differentially rotating disk are more difficult to stabilize; stability is achieved only for sufficiently large sound speed (although still of the order of  $c_T$ )

$$c_s \gtrsim c_{\text{crit}} \approx \frac{2\Omega}{\kappa} c_T \approx 2c_T \quad (17)$$

(Lin & Lau 1979; Morozov 1985; Griv et al. 2002). The modified Safronov-Toomre stability criterion given by Eq. (17) is obviously only an approximation, as it assumes spiral arm pitch angles  $\psi$  that are relatively small ( $|\tan\psi| \lesssim 1$ ) and neglects (among other things) the thickness of the disk. Nevertheless, simulations suggest that the formulation is approximately correct (Khoperskov et al. 2003; Liverts et al. 2003). As is seen, nonaxisymmetric gravity disturbances in a nonuniformly rotating system ( $d\Omega/dr \neq 0$ , or  $2\Omega/\kappa > 1$ ) are more difficult to suppress than the axisymmetric ones, in general agreement with the work by Goldreich & Lynden-Bell (1965) and Julian & Toomre (1966). The free kinetic energy associated with the



**Fig. 1.** A schematic model of a Jeans-unstable disk; **a)** the Safronov-Toomre unstable disk ( $Q < 1$ ) and **b)** the Safronov-Toomre stable disk ( $Q \geq 1$  but  $Q \lesssim 2\Omega/\kappa$ , and in spiral galaxies  $2\Omega/\kappa \approx 2$ ). In a differentially rotating disk ( $2\Omega/\kappa > 1$ ), nonaxisymmetric (spiral) modes of collective oscillations are more unstable than axisymmetric (radial) ones.

differential rotation of the system under study is one possible source for the growth of the energy of these spiral Jeans-type perturbations, and appears to be released when angular momentum is transferred outward.

The fact that self-gravitating, differentially rotating stellar disks can exhibit strong nonaxisymmetric responses even when the axisymmetric stability criterion is fulfilled, was most convincingly demonstrated already by Julian & Toomre (1966), concerning the gravitational effect of any single orbiting mass concentration (such as gas “lumps” in galaxies). These forced spiral waves are not to be confused, of course, with Lin & Shu’s fully self-consistent density wave proposal explored in the present theory. Similar result for self-gravitating gas disk was obtained by Goldreich & Lynden-Bell (1965). In Toomre (1981) this amplification was discussed in terms of “swing-mechanism”, very reminiscent of the way we reach the nonaxisymmetric stability criterion (17). Lau & Bertin (1978, p. 509) have clarified the problem by considering the motion of a fluid element: the density response that is in phase with the potential minimum is found to exceed, by an amount proportional to both  $d\Omega/dr$  and  $m$ , the corresponding response due to an axisymmetric field of equal strength.

Thus, if the disk is thin,  $c_s \ll r\Omega$ , and cold,  $c_s < c_T$  (or Toomre’s stability parameter  $Q < 1$ , respectively), then such a model will be unstable to both axisymmetric and nonaxisymmetric perturbations, and it should almost instantaneously (see below for a time estimate) taken on the form of a cartwheel, that is, a structure of spirals and rings (Fig. 1a). Clearly, in this case of axisymmetric and nonaxisymmetric excitation, the distribution of the surface density along the spiral arms is not uniform, but describes a sequence of maxima, that might be identified with forming gaseous complexes (Fig. A.1). Elmegreen & Elmegreen (1983) have argued that gravitational instability of the gas in spiral arms may be responsible for the formation of the observed gas complexes separated by distances of several kpc, which are regions of active star formation. The gas density in the disks of most galaxies is maintained at a level close to the threshold (marginal) stability of a gaseous layer to local gravitational perturbations (Zasov & Smirnova 2005). In this case, the regulation mechanism of the star formation rate associated with the growth of local gravitational instability in the layer must play a crucial role in the evolution of the gas

<sup>2</sup> At the limit of gravitational stability, the two conditions  $\partial\omega_J^2/\partial k = 0$  and  $\omega_J^2 \geq 0$  are fulfilled. The first condition determines the most unstable wavelength (the modified Jeans-Toomre wavelength)  $\lambda_{\text{crit}} \approx 2c_s^2/G\sigma_0$ , corresponding to the minimum on the dispersion curve (13). Use of the second condition determines the critical sound speed  $c_{\text{crit}}$  for the stability of arbitrary but not only axisymmetric perturbations.

content in the galactic disk. Such a Safronov-Toomre unstable disk should break up into discrete blobs of matter of preferred mass  $M_{\text{frag}}$  distributed in spirals around the spin axis<sup>3</sup>. These Jeans-unstable modes might be proposed as potential cloud-forming mechanism.

Contrary, if the disk is thin and *warm*,  $Q \geq 1$  but  $Q \lesssim 2\Omega/\kappa \approx 2$ , then such a model will be unstable only with respect to spiral perturbations (Fig. 1b) and cannot therefore fragment. An uncooled *hot* model with  $Q \gtrsim 2$  is Jeans-stable to all small perturbations, including the most unstable spiral ones. This is not an entirely new idea: Lau & Bertin (1978), Lin & Lau (1979), and Morozov (1985) first obtained the modified stability criterion  $c_{\text{crit}}$ . Simulated disks that prove stable evolve to  $Q$  values in the range 1.5–2.5 (Sellwood & Carlberg 1984; Athanassoula & Sellwood 1986; Tomley et al. 1991; Mihos et al. 1997; Bottema 2003). About the same value of  $Q \gtrsim 2$  brings also the observations of stellar disks (Toomre 1977; Bottema 1993). These values agree with modified stability criterion (17) that takes into account the fact that a nonuniformly rotating disk is more unstable to nonaxisymmetric perturbations. The instability is driven by a strong nonresonant interaction of the fluctuations (e.g., those produced by a spontaneous perturbation or, in rare cases, a satellite system) with the bulk of the particle population, and the dynamics of Jeans perturbations can be characterized as a nonresonant wave-particle interaction: in Eq. (7),  $\omega_* - l\kappa \neq 0$ .

The growth rate of the instability is relatively high,  $\Im\omega_* \approx \sqrt{2\pi G\sigma_0(k_s^2/|k|)}$ , and in general  $\Im\omega_* \sim \Omega$ , that is, the instability develops rapidly on a dynamical timescale (on a time of 2–3 disk rotations, or  $\lesssim 10^8$  yr in the spiral galaxy). An important feature of the instability under consideration is the fact that it is almost aperiodic ( $\Re\omega_*/\Im\omega_* \ll 1$ ). From Eq. (13), the growth rate of the instability has a maximum at the wavelength  $\lambda_{\text{crit}} \approx 2c_s^2/G\sigma_0$ . At the boundary of instability ( $Q \approx 1$ ),  $\lambda_{\text{crit}} \approx 2\pi^2 G\sigma_0/\kappa^2 \sim (2-4)\pi h$ ; thus  $\lambda_{\text{crit}} \gg h$ . It means that of all harmonics of initial gravity perturbation, one perturbation with  $\lambda_{\text{crit}} \approx 10h$ , with the associated number of spiral arms  $m_{\text{crit}}$ , and with the pitch angle  $\psi_{\text{crit}}$  will be formed asymptotically in time of a single disk rotation. Because  $\lambda_{\text{crit}} \gg h$ , the approximation of an infinitesimally thin disk used throughout the theory does not fail. Numerical experiments have already shown that in a steady, gravitoturbulent state, surface density structures have a similar characteristic physical scale  $\sim \lambda_{\text{crit}} \approx 60G\sigma_0/\Omega^2$  (the quantities  $c_s = 2c_T$  and  $\kappa = 1.5\Omega$  are used in calculating of the characteristic scale) that is independent of the size of the computational domain (Gammie 2001). In the solar neighborhood, the Safronov-Toomre critical sound speed  $c_T \sim 3 \text{ km s}^{-1}$  and the characteristic wavelength  $\lambda_{\text{crit}} \sim 2 \text{ kpc}$ . (These values for the gaseous disk are based on a surface density of  $6 M_\odot \text{ pc}^{-2}$  and an angular speed of  $25 \text{ km s}^{-1} \text{ kpc}^{-1}$ , e.g., Binney & Tremaine 1987). Interestingly, observations have indicated that in the solar vicinity of the Galaxy the distance between spiral arms  $\lambda_{\text{spir}} \approx 2 \text{ kpc}$  also. It is

<sup>3</sup> Axisymmetric and nonaxisymmetric features seen in Fig. 1a very reminiscent of the rings and spokes in the Cartwheel galaxy, the most striking example of the relatively small class of ring galaxies (Griv 2005).

natural to attribute the observed spiral structure of the Galaxy to the Jeans instability of gravity perturbations.

For those parameters of a disk ( $\lambda_{\text{crit}} \sim 1 \text{ kpc}$  and  $\sigma_0 \sim 1 M_\odot \text{ pc}^{-2}$ ), one obtains the typical mass of a gaseous complex  $M_{\text{cl}} \sim M_{\text{frag}} \sim (10^6-10^7) M_\odot$ . The latter is coincident in order of magnitude with the masses of giant molecular clouds, which have been discovered in the Galaxy (Binney & Tremaine 1987, p. 440).

In the another, opposite to (14) frequency range,

$$|\omega_*|^3 \sim 4\pi G\sigma_0\Omega(m/r|kL) \ll |\omega_J|^3, \quad (18)$$

that is,  $|\omega_*| \ll \Omega$ , Eq. (12) has another root equal to

$$\omega_{*3} \approx 4\pi G\sigma_0(\Omega/\omega_J^2)(m/r|kL). \quad (19)$$

The root (19) describes the gradient ( $L^{-1} \neq 0$ ) branch of oscillations. As is seen, the gradient perturbations are stable and are independent of the stability of Jeans modes. These stable ( $\Im\omega_{*3} = 0$ ) oscillations are obviously not important in dynamics of galaxies. Analogous oscillation branches, with frequencies proportional to the gradients of the undisturbed quantities, occur in plasmas (drift waves) and in the terrestrial oceans (Rossby waves).

The Jeans instability of a disk will in fact be suppressed because of the efficient heating of the medium during the growth of that same instability (Morozov 1978; Fuchs 2001; Griv et al. 2002; Appendix B of this paper). Following the physics of plasmas, by heating of galactic disks we understand the transfer of ordered energy of rotation or the energy of discrete oscillations, into energy of random motion and ultimately into heat, due to turbulence, i.e., the nonlinear interaction between oscillations.

#### 4. Angular momentum transfer

The collective torque per unit area in the galaxy exerted by the gravity perturbations on the disk is

$$\frac{1}{r} \frac{d\Gamma}{dr} = - \left\langle \int_0^{2\pi} \sigma_1(r, \varphi') \frac{\partial \Phi_1(r, \varphi')}{\partial \varphi'} d\varphi' \right\rangle. \quad (20)$$

Using Eq. (7), in terms of the Fourier components defined in Eq. (1),  $\Gamma = \sum_{m=1}^{\infty} \Gamma_m$ , from Eq. (20) one finds

$$\frac{1}{r} \frac{d\Gamma_m}{dr} \approx -8\pi \frac{m^2}{rL} \frac{\sigma_0\Omega}{\kappa^2 \Im\omega_*} \Phi_1 \Phi_1^* \quad \text{if } \Im\omega_* > 0, \quad (21)$$

or  $(1/r)(d\Gamma_m/dr) = 0$  if  $\Im\omega_* \leq 0$ ,  $\Phi_1 \Phi_1^* = |\tilde{\Phi}|^2 \exp(2\Im\omega_* t)$ , and  $\langle \dots \rangle$  denotes the time average over the fast oscillations. Equation (21) is correct only in the main domain of the system under study between the inner and outer Lindblad resonances. A special analysis of the solution near corotation ( $\omega_* = 0$ ) and Lindblad ( $\omega_* \pm \kappa = 0$ ) resonances is required. Thus, the points  $r_{\text{ILR}}$  and  $r_{\text{OLR}}$  in which  $\omega_* \pm \kappa = 0$  are called the points of inner and outer Lindblad resonances. They play an important role in the theory: the solution of spiral type given by Eq. (1) rapidly oscillating in the radial direction ( $|k_r|r \gg 1$ ) lies between  $r_{\text{ILR}}$  and  $r_{\text{OLR}}$ . Outside the resonances,  $r < r_{\text{ILR}}$  and  $r > r_{\text{OLR}}$ , the solution decreases exponentially (Lin & Lau 1979). Resonances of a higher order are dynamically of less importance (Lin et al. 1969). To emphasize it again, the present analysis is restricted

to consideration of the main part of a disk between the Lindblad resonances. The spatially limited resonances have been investigated by Lynden-Bell & Kalnajs (1972), Goldreich & Tremaine (1979), and Griv et al. (2000).

Four physical conclusions can be deduced from Eq. (21). (a) The distribution of the angular momentum of a disk will be changed under the action of only the *nonaxisymmetric* forces  $\propto m$ . The latter is obvious: axially symmetric motions of a system produce no gravitational couplings between the inner parts and the outer parts. (b) The distribution of the angular momentum will be changed with time only under the action of *growing*, that is, gravitationally (Jeans-) *unstable* perturbations ( $\Im\omega_* > 0$ ). (c) Unstable spiral perturbations can transfer angular momentum only in spatially *inhomogeneous* disks ( $(\partial/\partial r)(\sigma_0\Omega\kappa^{-2}) \neq 0$ ). This result is also anticipated, because in homogeneous disks the angular velocity of spiral perturbations  $\Omega_p = 0$  (Eq. (16)), and therefore there is no exchange of angular momentum in the wave-gas system. And finally (d) because in gravitating disks in equilibrium  $L < 0$  (see Appendix C),  $(1/r)(d\Gamma_m/dr) > 0$ : an applied gravitational torque *increases* the angular momentum of the gas element and thus leads to motion of the gas element at a larger radius and thus tends to decrease  $\Omega$ . (As is known,  $\Omega(r)$  in a self-gravitating disk is a decreasing function of  $r$ , whereas the angular momentum of a unit mass,  $\Omega r^2$ , is an increasing function of  $r$ .) This takes place in the main domain of the disk between the Lindblad resonances where spiral density waves are self-excited via a non-resonant wave-fluid interaction. This in turn cannot be done for all masses because the total orbital momentum must remain constant. As a result, a group of inner particles with radii  $r < r_{\text{ILR}}$  moves inward and a group of particles with radii  $r_{\text{ILR}} < r < r_{\text{OLR}}$  moves outward. The secular evolution of disk galaxies, therefore, proceeds in the direction of increasing central mass concentration in the baryonic material and of extending of outer portions. As a result, the bulk of angular momentum is transferred radially outward whereas an inner gaseous medium moves inward, losing a part of the angular momentum. (Additional emission and absorption of angular momentum occur at resonances (Lynden-Bell & Kalnajs 1972).) Lynden-Bell & Kalnajs (1972) have proved that the torques can only communicate angular momentum outward if the spirals trail, thus explaining why trailing spiral structure is prevalent in galaxies. Lynden-Bell & Kalnajs (and Goldreich & Tremaine 1979) considered, however, only wave-particle interactions at resonances in stable systems. The outward transfer of orbital momentum allows the central parts of galaxies to contract, and the remnant disk of gas is the reservoir for forming clouds.

Equation (21) derived in this paper to describe the wave-gas exchange of angular momentum for unstable disks is a new result, and one has to realize that this is the focus of the paper. At the qualitative level, the influence of viscosity on a rotating mass of gas has been well understood (Lynden-Bell & Pringle 1974; Pringle 1981). By viscosity we mean the mechanism, whatever it is, which enables angular momentum to be transferred and energy to be dissipated. Disk viscosity would regulate evolution and structure formation of disk galaxies. In general, “the dissipative processes act to spread the disk out,

allowing the inner parts to move in and necessitating, through conservation of angular momentum, the outer parts to move out” (Pringle 1981, p. 140). Lynden-Bell & Kalnajs (1972) have also suggested that gravitating disks want to transfer their angular momentum outwards. Simulations have already demonstrated that instabilities produce growing spiral waves and associated gravitational torques, which are potent agents of angular momentum transport.

Equation (21) shows that although the angular momentum exchange between gas and the growing wave is of second order in the potential  $\Phi_1$ , it is obtained by taking inner products of first order solutions. Thus, going to nonlinear approximation of the theory is unnecessary to derive the nonresonant wave-gas angular momentum exchange (cf. Lynden-Bell & Kalnajs 1972, p. 10).

According to Eq. (21), the angular momentum transfer efficiency of nonaxisymmetric Jeans-unstable density waves depends on their spatial and temporal form. Let us evaluate the collective torque for a realistic model of the disk. In accordance with the theory, the fastest growing mode with  $m \approx 1$ ,  $k_* = k_{\text{crit}}$ , and  $\Im\omega_* \sim \Omega$  is considered. Taking into account that  $8\pi m^2\Phi_1\Phi_1^* \sim \Phi_0^2$  (an astrophysicist might well consider a perturbation with  $\Phi_1/\Phi_0$  of 1/10 to be quite small) and  $\Phi_0 \sim r^2\Omega^2/2$ , where  $\Phi_0$  is the basic potential, from Eq. (21) one obtains  $|(1/r)(d\Gamma_m/dr)| \sim \sigma_0 r^3 \Omega^2 / 4|L|$ . The angular momentum of the given gaseous element  $\mathcal{P} = \sigma_0 r^2 \Omega$ . Then the characteristic time of the angular momentum redistribution is  $t \sim \mathcal{P} / |(1/r)(d\Gamma_m/dr)| \sim (4|L|/r)\Omega^{-1} \sim \Omega^{-1}$ . Thus, already in the first disk revolution, in, say, about  $10^8$  yr, the Jeans-unstable galactic disk sees a large portion of its angular momentum transferred outward, and mass transferred both inward and outward (see Fig. D.1).

Paczynski (1978) considered a model of a thin disk that is marginally stable to self-gravity. His model followed the accretion-disk theory by Shakura & Sunyaev (1973) and Lynden-Bell & Pringle (1974). These standard models have “understood” angular momentum transport in that they assumed that it is due to turbulent stresses. Accordingly, “turbulent viscosity” driven by local disk instabilities was assumed to transfer angular momentum outwards and makes the accretion of matter onto the central object possible. There is one important difference between our model and the models by Paczynski (1978), Shakura & Sunyaev (1973), Lynden-Bell & Pringle (1974): in our model the physical cause for the turbulence is included self-consistently by considering a self-consistent system of the gasdynamic and Poisson equations.

## 5. Discussion

In this paper, inhomogeneity and differential rotation are taken into proper account in the study of instabilities of a gravitating disk. It is shown that the Jeans instability may be an effective generating mechanism for the spiral  $\sim 1$  kpc structure of a gaseous medium of a flat galaxy, developing in the main part of the system between the inner and outer Lindblad resonances. Axisymmetric Safronov-Toomre instability breaks up the spiral arms into discrete gaseous complexes of a mass of  $(10^6 - 10^7) M_\odot$  each. Spiral density perturbations arising from

nonresonant Jeans instabilities grow almost aperiodically and exert torques which effectively redistribute both the mass and the angular momentum of a disk, giving the disk a source of internal viscosity. The latter may provide the required accretion rate. The spiral modes are maintained by an outward transport of angular momentum in an inhomogeneous disk. In the quasi-steady state after a time of only 2–3 rotations of the system, a fraction of the mass (eventually residing in the central parts) retains only a minor fraction of the angular momentum, most of the latter deposited in outer portions of a disk. This leads to an exponential-like mass density profile in disk galaxies. Interestingly, one of the most notable properties of flat galaxies is that the mass profiles of their both stellar and gaseous disks are remarkably close to similar exponential-like distributions over a large dynamic range in radii (Freeman 1970; Meijerink et al. 2005). Redistribution of the gas in a disk under the influence of a growing spiral potential may be also crucial to the understanding of the fuelling mechanism of AGNs.

In addition to the radial spreading, this mechanism increases the effective temperature of the disk. Sufficient temperature (the sound speed  $c_s \gtrsim (2\Omega/\kappa)c_T \approx 2c_T$ , or Toomre’s stability parameter  $Q = 2\text{--}2.5$ , respectively) and core-dominated exponential-like mass density distribution prevent the Jeans instability from occurring. The secular evolution of disk galaxies, therefore, proceeds in the direction of increasing central mass concentration in the baryonic material, of extending outer portions, and of dynamical heating of the galactic material (see also Griv et al. 2002 and Griv & Gedalin 2004 for a discussion). The diffusion in configuration space is due entirely to the almost aperiodic (in a rotating frame) growth of the nonresonant Jeans-unstable disturbances in a self-gravitating system subject to a time-dependent potential.

Thus, as a result of the gravitational instability the gaseous disk of a galaxy is changing in the plane rapidly during a time of about 2–3 rotations ( $\sim 10^8$  yr) only, and therefore the spiral structures are transient. Fresh waves must be continually created to maintain the spiral pattern. One obvious way to generate long-lived density wave structures is to cool the disk by radiation and thus provoke Jeans instabilities. Even optically thick Jeans-stable gaseous disk will eventually cool down because of the radiative heat losses, and its thickness will decrease. As a result local gravitational instabilities will be turned on again, and we have completed the cycle. Obviously, in this cycle some thermal energy is radiated away. Gravitationally unstable *cooled* disks evolve through the continued action of fresh Jeans-unstable density waves (cf. Tomley et al. 1991 and Gammie 2001); the cold interstellar medium may play a dominant role in determining the long-term wave-like structure in galaxies. In this regard, no prominent spirals are seen in SO galaxies that have little or no interstellar matter. Note that the same problem arises with spiral structures in the plane of a stellar disk of a galaxy in  $N$ -body codes since they often show that after forming spirals the disk becomes hot enough,  $Q \gtrsim 2$ , and then – unless some physics is added to continuously cool it – as the result of this heating spirals dissipate after a few rotations of the system (Miller et al. 1970; Hohl 1972; Quirk 1972; Sellwood & Carlberg 1984; Griv & Chiueh 1998). Apparently, Goldreich & Lynden-Bell (1965) were the

first to propose that the amplitude of gravitational instabilities, and hence the strength of their influence on the disk, is directly regulated by the balance between cooling and heating processes in the disk. This idea was subsequently developed by Paczynski (1978) and Lin & Pringle (1990), and tested in numerical simulations by Tomley et al. (1991), Gammie (2001), and others.

The cold disk with  $Q < 1$  is subject to both radial and spiral instabilities and might therefore be clumpy (Fig. 1a). On the other hand, the warm uncooled disk with  $1 \leq Q \lesssim 2$  is subject to only spiral instabilities (Fig. 1b) and cannot therefore fragment. The hot core-dominated uncooled disk with  $Q > 2$  is stable against all perturbations, including the most unstable spiral ones.

Because the typical relaxation time (the time required to stabilize the initially unstable disk)  $t \sim (\mathfrak{I}\omega_*)^{-1} \sim \Omega^{-1}$ , the disk fragments when  $\langle\tau\rangle\Omega \lesssim 1$ , where  $\langle\tau\rangle$  is an effective cooling time defined as the average internal energy of the model divided by average cooling rate (Gammie 2001). Thus, I believe to have obtained a theoretical interpretation of Johnson & Gammie’s (2003) experimental result: gaseous disks with  $\langle\tau\rangle\Omega \lesssim 1$  fragment on the dynamical timescale  $\lesssim \Omega^{-1}$  and thus do not exist.

A cooled disk could generate transient Jeans-unstable density waves even without an external trigger (Miller et al. 1970; Sellwood & Carlberg 1984; Griv & Chiueh 1998). Note especially the work by Quirk (1972), who has shown that the tangential gravitational forces produced by the spiral arms are necessary to maintain the long-term recurrent density wave structure of the galaxy in the “gas clouds” component, which was *dissipative* through inelastic cloud-cloud collisions. The computer generated long-term patterns look like real galaxies in that spiral patterns appear in cooled “gas” while the uncooled “stars” form a background that shows much less structure. As cooling process always exists in the disk galaxies, the Jeans instability can be considered to be a long-term generating mechanism for unstable density waves, thereby leading to recurrent short-lived  $\sim 5 \times 10^8$  yr spiral or spiral-ring patterns in the gas-rich flat galaxies (see Fig. 1).

## References

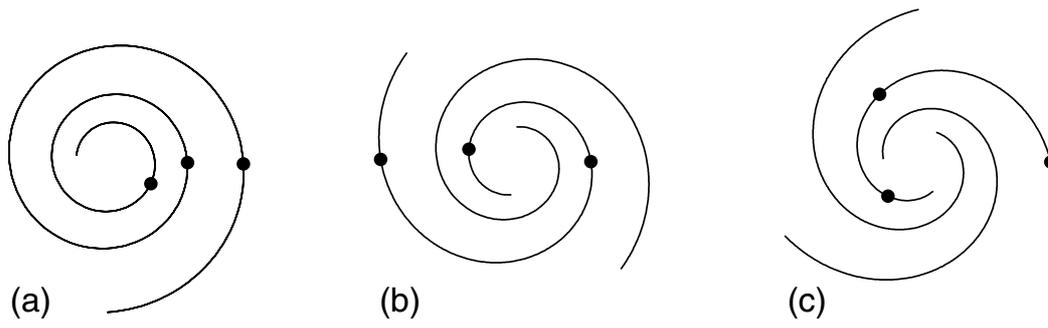
- Athanassoula, E., & Sellwood, J. A. 1986, MNRAS, 221, 213
- Berman, R. H., & Mark, J. W.-K. 1977, ApJ, 216, 257
- Bertin, G., Lin, C. C., Lowe, S. A., & Thurstans, R. P. 1989, ApJ, 338, 104
- Bertin, G., & Mark, J. W.-K. 1978, A&A, 64, 389
- Binney, J., & Tremaine, S. 1987, Galactic Dynamics (Princeton: Princeton Univ. Press)
- Bottema, R. 1993, A&A, 275, 16
- Bottema, R. 2003, MNRAS, 344, 358
- Elmegreen, B. G., & Elmegreen, D. M. 1983, MNRAS, 203, 31
- Freeman, K. C. 1970, ApJ, 160, 811 (erratum 161, 802)
- Fuchs, B. 2001, MNRAS, 325, 1637
- Gammie, C. F. 2001, ApJ, 553, 174
- Goldreich, P., & Lynden-Bell, D. 1965, MNRAS, 130, 125
- Goldreich, P., & Tremaine, S. 1979, ApJ, 233, 857
- Griv, E. 2005, Ap&SS, 299, 371
- Griv, E., & Chiueh, T. 1998, ApJ, 503, 186

- Griv, E., & Gedalin, M. 2004, *AJ*, 128, 1965
- Griv, E., Gedalin, M., Eichler, D., & Yuan, C. 2000, *Phys. Rev. Lett.*, 84, 4280
- Griv, E., Yuan, C., & Gedalin, M. 1999, *MNRAS*, 307, 1
- Griv, E., Gedalin, M., & Yuan, C. 2002, *A&A*, 383, 338
- Haywood, M., Robin, A. C., & Crézé, M. 1997, *A&A*, 320, 440
- Hohl, F. 1972, *J. Comput. Phys.*, 9, 10
- Hunter, C. 1965, *MNRAS*, 129, 321
- Johnson, B. M., & Gammie, C. F. 2003, *ApJ*, 597, 131
- Julian, W. H., & Toomre, A. 1966, *ApJ*, 146, 810
- Khoperskov, A. V., Zasov, A. V., & Tyurina, N. V. 2003, *Astron. Rep.*, 47, 357
- Krall, N. A., & Trivelpiece, A. W. 1986, *Principles of Plasma Physics* (repr.; San Francisco: San Francisco Press)
- Landau, L. D., & Lifshitz, E. M. 1987, *Fluid Mechanics* (Oxford: Pergamon)
- Lau, Y. Y., & Bertin, G. 1978, *ApJ*, 226, 508
- Laughlin, G., Korchagin, V., & Adams, F. C. 1998, *ApJ*, 504, 945
- Laughlin, G., & Rózyczka, M. 1996, *ApJ*, 456, 279
- Lifshitz, E. M., & Pitaevskii, L. P. 1981, *Physical Kinetics* (Oxford: Pergamon)
- Lin, C. C., & Lau, Y. Y. 1979, *SIAM Stud. Appl. Math.*, 60, 97
- Lin, C. C., Yuan, C., & Shu, F. H. 1969, *ApJ*, 155, 721
- Lin, D. N. C., & Pringle, J. E. 1990, *ApJ*, 358, 515
- Liverts, E., Griv, E., Gedalin, M., & Eichler, D. 2003, in *Galaxies and Chaos*, ed. G. Contopoulos, & N. Voglis (Berlin: Springer), 340
- Lovelace, R. V. E., & Hohlfeld, R. G. 1978, *ApJ*, 221, 51
- Lynden-Bell, D., & Kalnajs, A. J. 1972, *MNRAS*, 157, 1
- Lynden-Bell, D., & Pringle, J. E. 1974, *MNRAS*, 168, 603
- Mark, J. M.-K. 1977, *ApJ*, 212, 645
- Mayer, L., Quinn, T., Wadsley, J., & Stadel, J. 2004, *ApJ*, 609, 1042
- Meijerink, R., Tilanus, R. P. J., Dullemond, C. P., Israel, F. P., & an der Werf, P. P. 2005, *A&A*, 430, 427
- Mihos, J. C., McGaugh, S. S., & de Blok, W. J. G. 1997, *ApJ*, 477, 79
- Miller, R. H., Prendergast, K. H., & Quirk, W. J. 1970, *ApJ*, 161, 903
- Montenegro, L. E., Yuan, C., & Elmegreen, B. G. 1999, *ApJ*, 520, 592
- Morozov, A. G. 1978, *Soviet Ast. Lett.*, 4, 115
- Morozov, A. G. 1981, *Soviet Ast. Lett.*, 7, 109
- Morozov, A. G. 1985, *Soviet Ast.*, 29, 120
- Paczynski, B. 1978, *Acta Astr.*, 28, 91
- Papaloizou, J. C., & Lin, D. N. C. 1995, *ARA&A*, 33, 505
- Pickett, B. K., Mejía, A. C., Durisen, R. H., Cassen, P. M., Berry, D. K., & Link, R. P. 2003, *ApJ*, 590, 1060
- Pringle, J. E. 1981, *ARA&A*, 19, 137
- Quirk, W. J. 1972, in *Gravitational N-Body Problem*, IAU Coll. 10, ed. M. Lecar (Dordrecht: Reidel), 250
- Sagdeev, R. Z., & Galeev, A. A. 1969, *Nonlinear Plasma Theory*, ed. T. M. O'Neil, & D. L. Book (New York: Benjamin)
- Safronov, V. S. 1960, *Ann. d'Astr.*, 23, 979
- Sellwood, J. A., & Carlberg, R. G. 1984, *ApJ*, 282, 61
- Shakura, N. I., & Sunyaev, R. A. 1973, *A&A*, 24, 337
- Shu, F. H. 1970, *ApJ*, 160, 99
- Shu, F. H. 1984, in *Planetary Rings*, ed. R. Greenberg, & A. Brahic (Tucson: Univ. of Arizona Press), 513
- Tomley, L., Cassen, P., & Steiman-Cameron, T. 1991, *ApJ*, 382, 530
- Toomre, A. 1964, *ApJ*, 139, 1217
- Toomre, A. 1977, *ARA&A*, 15, 437
- Toomre, A. 1981, in *Structure and Evolution of Normal Galaxies*, ed. S. M. Fall, & D. Lynden-Bell (Cambridge: Cambridge Univ. Press), 111
- Zasov, A. V., & Smirnova, A. A. 2005, *Astron. Lett.*, 31, 160

E. Griv: Gravitationally unstable gaseous disks of flat galaxies, *Online Material p 1*

# Online Material

**Appendix A: Maxima perturbed density**



**Fig. A.1.** **a)** Gravitationally unstable density waves with  $m = 1$  arm in the  $(r, \varphi)$ -plane, **b)** density waves with  $m = 2$  arms, and **c)** density waves with  $m = 3$  arms. The filled circles represent the maxima of the perturbed density of Jeans waves, which are unstable to both axisymmetric and nonaxisymmetric gravity perturbations (Toomre's  $Q$ -value under 1, or the value of the sound speed  $c_s$  smaller than  $c_T$ , respectively). The distribution of the surface density along the spiral arms is not uniform, but describes a sequence of maxima, that might be identified with forming gaseous complexes (gaseous clouds). The preferred mass of each fragment (cloud) is about  $M_{\text{frag}} \sim 4c_s^4/G^2\sigma_0$ .

## Appendix B: The dynamical heating

Gravitational instabilities convert potential energy into kinetic energy of random motions. This process—the self-suppression of instability by a rise in the amplitude of the unstable harmonics—recalls the case of nonresonant quasilinear relaxation in plasmas, which can effectively heat the medium without raising the entropy (Sagdeev & Galeev 1969; Krall & Trivelpiece 1986, p. 531). Indeed, the perturbed potential energy  $U$  is a negative quantity:

$$U = \left\langle \frac{\sigma_1 \Phi_1^*}{2} \right\rangle = -\frac{|k| |\tilde{\Phi}|^2}{2\pi G} e^{2\Im\omega_* t}, \quad (\text{B.1})$$

where  $\langle \cdot \rangle$  denotes the time average over the fast oscillations,  $\sigma_0 = \langle \sigma \rangle$ ,  $\Phi_0 = \langle \Phi \rangle$ ,  $\langle \sigma_1 \rangle = \langle \Phi_1 \rangle = 0$ , and  $\Phi_1^*$  is the complex conjugate potential. Clearly as gravitational instability develops the amplitudes  $|U|$  of the unstable Fourier-harmonics of the perturbation will grow (via the mass and orbital momentum redistribution), setting energy free to heat the medium. This effect will in turn diminish the growth rate of the instability and ultimately cause it to become saturated. Fresh waves must be continually created to maintain the density wave pattern. Equation (B.1) shows that random motions of gas elements of an unstable disk will grow under the action of both the radial and the tangential forces ( $k \equiv \sqrt{k_r^2 + k_\varphi^2} \neq 0$ ). Of course, this is not really temperature, it is ordered motion, i.e., coherent mechanical scillations of the gas element in response to the fluctuating fields. Numerical experiments have already confirmed that uncooled gravitationally unstable disks are heated to the point of stability and cease to produce density waves in a time of a single rotation (Hohl 1972; Sellwood & Carlberg 1984; Athanassoula & Sellwood 1986; Tomley et al. 1991; Laughlin & Różyczka 1996; Liverts et al. 2003).

Following Morozov (1978), let us obtain mathematical expressions describing the heating process. For simplicity, the gas is supposed to obey a simple polytropic law

$$P = P_0 (\sigma / \sigma_0)^n, \quad (\text{B.2})$$

where the two-dimensional adiabatic index  $n$  can be mapped to a three-dimensional adiabatic index in the low-frequency

(static) limit; in astrophysical disks  $n = 1.5 - 2$ . Equation (B.2) can be expanded as far as terms quadratic in the amplitude of the perturbed density

$$P = P_0 + n \frac{P_0}{\sigma_0} \delta\sigma + \frac{n(n-1)}{2} \frac{P_0}{\sigma_0^2} \delta\sigma^2, \quad (\text{B.3})$$

where  $\delta\sigma = \epsilon\sigma_1 + \epsilon^2\sigma_2$  and  $\epsilon \ll 1$ . Next, the result can be averaged in the sense indicated above:

$$\langle P - P_0 \rangle = \delta P(t) = \frac{n(n-1)P_0}{\sigma_0^2} |\tilde{\sigma}|^2 e^{2\Im\omega_* t}, \quad (\text{B.4})$$

where the growth rate  $\Im\omega_*(t)$  is determined from Eq. (13) by the relation

$$\Im\omega_* = \left[ -\kappa^2 + 2\pi G \sigma_0 (k_*^2 / |k|) - n(P_0 + \delta P(t)) k_*^2 / \sigma_0 \right]^{1/2}. \quad (\text{B.5})$$

Equations (B.4) and (B.5) completely describes the heating process. Of course, the system of Eqs. (B.4)–(B.5) is correct only in the linear approximation used throughout the theory,  $|\sigma_1 / \sigma_0| \ll 1$  and  $|\Phi_1 / \Phi_0| \ll 1$ . The linear wave considered in the present study is growing on the timescale of about one cycle rotation period ( $\sim \Omega^{-1}$ ) and in one cycle rotation period it will reach nonlinear amplitudes. Strictly speaking, this theory of weak turbulence describes only the tendency of the disk to be heated by Jeans-unstable density waves and shows the direction of the disk's evolution. One understands that we still have to develop the theory of strong turbulence.

Taking into account the fact that  $c_s^2 \propto \langle P - P_0 \rangle$ , from Eq. (B.4) one obtains that during the quasilinear stage of the instability the effective temperature  $c_s^2$  grows with time according to the law

$$c_s^2 \propto t. \quad (\text{B.6})$$

Of course, this is not really temperature, it is ordered motion, i.e., coherent mechanical oscillations of the gas element in response to the fluctuating fields. Interestingly, observations convincingly indicate the same increase in the velocity spread in later-type main-sequence stars of the solar vicinity:

$$c_r \propto \sqrt{t},$$

where  $c_r$  is the radial dispersion of peculiar (random) velocities and  $t$  is an age of stars (e.g., Binney & Tremaine 1987, p. 470; Griv et al. 2002). HIPPARCOS data have also implied significant heating of the Galactic disk by moderately tightly wound spiral structures; the Galactic potential is significantly nonaxisymmetric at the solar radius. The steady growth in  $c_r \propto \sqrt{t}$  has already been detected in an unstable  $N$ -body disk (Sellwood & Carlberg 1984; Binney & Tremaine 1987, p. 480; Liverts et al. 2003).

It is obvious that the uncooled disk manages to keep its local stability parameter close to the critical value,  $Q \approx 2$  (Griv et al. 1999, 2002; Griv & Gedalin 2004). In this case, once the differentially rotating disk has been heated to values  $Q \approx 2$  (or  $c_s \approx 2c_T$ , respectively) by gravitationally unstable density waves, no further unstable waves can be sustained by virtue of the Jeans instability—unless some cooling mechanism is available leading to Toomre's  $Q$ -value under approximately 2 or to the value of  $c_s$  smaller than approximately  $2c_T$ , respectively. To repeat myself, it has been found

in simulations that the stability number  $Q$  of Toomre in relaxed equilibrium disks does not fall below a critical value, which lies about  $Q_{\text{crit}} = 2\text{--}2.5$  (Hohl 1972; Sellwood & Carlberg 1984; Athanassoula & Sellwood 1986; Tomley 1991; Mihos et al. 1997; Bottema 2003). However, no adequate explanation of the latter has been presented (see also Morozov 1981; Griv & Chiueh 1998; Liverts et al. 2003; and Khoperskov et al. 2003 for a discussion).

The development of gravitational instabilities would lead to a self-regulation process. Namely, if the differentially rotating disk is initially cold ( $Q < Q_{\text{crit}}$ ), then gravitational instabilities would repeatedly heat it according to the law (B.6) on the dynamical timescale  $\lesssim \Omega^{-1}$ , bringing it toward stability. On the other hand, if the disk is initially hot enough ( $Q \geq Q_{\text{crit}}$ ), then radiative cooling is going to bring the value of Toomre's stability parameter  $Q$  down toward an unstable configuration. A similar approach has been suggested in computer simulations (e.g., Gammie 2001).

### Appendix C: The distribution function for angular momentum

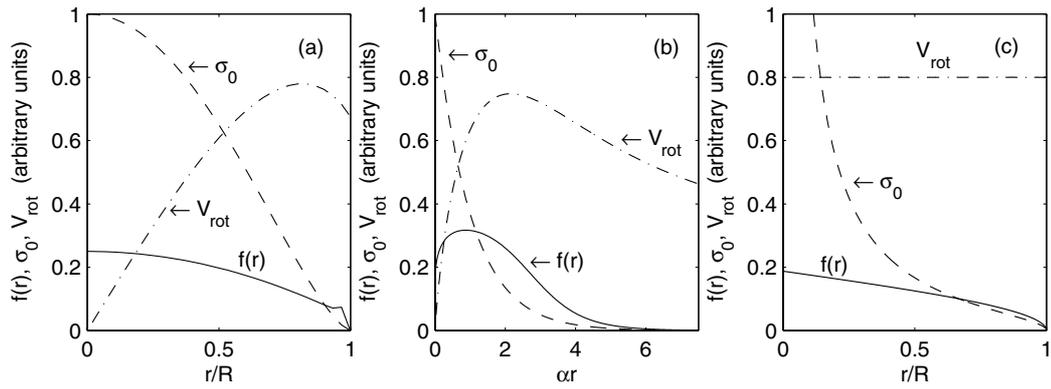
The time behavior of the quantity  $\sigma_0 \Omega \kappa^{-2}$  in self-gravitating disks has been considered by Berman & Mark (1977). They have proved that in the axisymmetric case

$$\frac{d}{dt} \left( \frac{\sigma_0 \Omega}{\kappa^2} \right) = 0, \quad (\text{C.1})$$

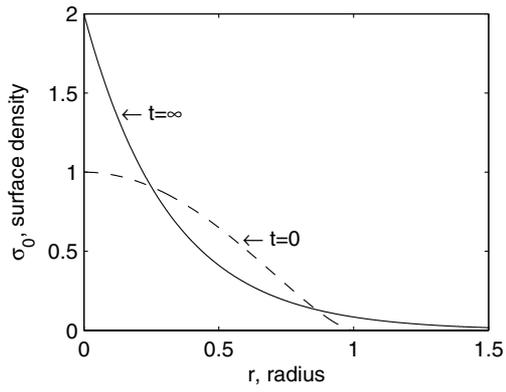
that is, vorticity is carried along with the fluid. Equation (C.1) is a consequence of the detailed conservation of mass and angular momentum (in the axisymmetric case). Lovelace & Hohlfield (1978) have pointed out that

$$f(r) = \sigma_0 \Omega \kappa^{-2}$$

has the role of the distribution function for angular momentum. In self-gravitating disks in equilibrium,  $f(r)$  is a decreasing function over a range of  $r$ ; thus, in main parts of flat galaxies  $L < 0$ . Values of  $\Omega(r)$ ,  $\kappa(r)$ , and  $\sigma_0(r)$  derived from the oftenly used equilibrium models of galaxies, namely Hunter's (1965)  $n = 1$  model, exponential disk model (Binney & Tremaine 1987, p. 78), and Hunter's  $n \rightarrow \infty$  model are used to calculate  $f(r)$  for the three cases. The  $f(r)$  curves are shown in Fig. C.1. Accordingly, in all cases  $f(r)$  is a decreasing function of  $r$  in the main parts of the systems. In the solar vicinity of the Galaxy the value of the radial scale length  $|L|$  is about 3–4 kpc, which is a typical value for the radial scale length of the exponential component of the disk when compared with external galaxies of similar morphological types.



**Fig. C.1.** The distribution function for angular momentum  $f(r) = \sigma_0 \Omega \kappa^{-2}$  (Lovelace & Hohlfield 1978, p. 55) vs. a normalized radius  $r/R$  (or  $\alpha r$ ) for the equilibrium galactic models calculated from the surface density  $\sigma_0(r)$ , the angular rotation velocity  $\Omega(r) = V_{\text{rot}}/r$ , and the epicyclic frequency  $\kappa(r)$ ; **a)** Hunter's  $n = 1$  model, **b)** the exponential disk model, and **c)** Hunter's  $n \rightarrow \infty$  model. In main parts of self-gravitating disks in equilibrium  $f(r)$  is a decreasing function of a radial distance  $r$ .



**Fig. D.1.** Surface mass density of a disk  $\sigma_0(r) = \langle \sigma(r) \rangle$  before ( $t = 0$ ) and after ( $t = \infty$ ) initially Jeans-unstable gravity disturbances have increased. Gravitationally (Jeans-) unstable nonaxisymmetric ( $m \neq 0$ ) perturbations transfer angular momentum outward, and mass both inward and outward in spatially inhomogeneous disks.

#### Appendix D: Evolution of the surface mass density